

Andy Pavlo

Carnegie Mellon University
Department of Computer Science
Gates-Hillman Center 9019
Pittsburgh, PA 15213-3891 USA

E-mail: pavlo@cs.cmu.edu
Web: <https://cs.cmu.edu/~pavlo/>
Mastodon: @andy_pavlo@discuss.systems
ORCID: 0000-0001-6040-6991

Education

Ph.D., Computer Science Brown University	Summer 2013 <i>Providence, RI USA</i>
M.Sc., Computer Science Brown University	Spring 2009 <i>Providence, RI USA</i>
M.Sc., Computer Science Rochester Institute of Technology	Spring 2006 <i>Rochester, NY USA</i>
B.Sc. with Honors, Computer Science Rochester Institute of Technology	Spring 2005 <i>Rochester, NY USA</i>

Appointments

Associate Professor with Indefinite Tenure Carnegie Mellon University, Department of Computer Science	July 2022 – present <i>Pittsburgh, PA USA</i>
Associate Professor Carnegie Mellon University, Department of Computer Science	July 2019 – June 2022 <i>Pittsburgh, PA USA</i>
Assistant Professor Carnegie Mellon University, Department of Computer Science	September 2013 – June 2019 <i>Pittsburgh, PA USA</i>

Professional Experience

Advisory Board Voltron Data, Inc.	January 2021 – present <i>San Francisco, CA USA</i>
Co-Founder and CEO OtterTune, Inc.	May 2020 – August 2024 <i>Pittsburgh, PA USA</i>
Advisory Board Fauna, Inc.	April 2019 – April 2021 <i>San Francisco, CA USA</i>
Advisory Board Splice Machine, Inc.	November 2018 – August 2021 <i>San Francisco, CA USA</i>
Expert Witness Much Shelist, P.C.	May 2014 – May 2016 <i>Chicago, IL USA</i>
Research Assistant Brown University, Department of Computer Science	September 2007 – August 2013 <i>Providence, RI USA</i>
Non-Testifying Technical Expert Choate, Hall & Stewart, LLP.	June 2009 – November 2011 <i>Boston, MA USA</i>
Research Intern Vertica, Inc.	June 2008 – August 2008 <i>Billerica, MA USA</i>
Systems Programmer – Condor Project University of Wisconsin–Madison, Department of Computer Sciences	September 2005 – August 2007 <i>Madison, WI USA</i>

Teaching Experience

- **Advanced Database Systems** (Carnegie Mellon University, 15-721) – Spring 2024
- **Intro to Database Systems** (Carnegie Mellon University, 15-445/645) – Fall 2023
- **Advanced Database Systems** (Carnegie Mellon University, 15-721) – Spring 2023
- **Intro to Database Systems** (Carnegie Mellon University, 15-445/645) – Fall 2022
- **Special Topics: Self-Driving Database Systems** (Carnegie Mellon University, 15-799) – Spring 2022
- **Intro to Database Systems** (Carnegie Mellon University, 15-445/645) – Fall 2020
- **Advanced Database Systems** (Carnegie Mellon University, 15-721) – Spring 2020
- **Intro to Database Systems** (Carnegie Mellon University, 15-445/645) – Fall 2019
- **Advanced Database Systems** (Carnegie Mellon University, 15-721) – Spring 2019
- **Intro to Database Systems** (Carnegie Mellon University, 15-445/645) – Fall 2018
- **Advanced Database Systems** (Carnegie Mellon University, 15-721) – Spring 2018
- **Intro to Database Systems** (Carnegie Mellon University, 15-445/645) – Fall 2017
- **Advanced Database Systems** (Carnegie Mellon University, 15-721) – Spring 2017
- **Database Applications** (Carnegie Mellon University, 15-415/615) – Fall 2016
- **Advanced Database Systems** (Carnegie Mellon University, 15-721) – Spring 2016
- **Database Applications** (Carnegie Mellon University, 15-415/615) – Fall 2015
- **Database Applications** (Carnegie Mellon University, 15-415/615) – Spring 2015
- **Database Applications** (Carnegie Mellon University, 15-415/615) – Spring 2014
- **Special Topics: Big Data Systems** (Carnegie Mellon University, 15-799) – Fall 2013
- **NewSQL – The Course** (Brown University, CSC2270) – Spring 2012
- **Non-standard, Web-scale Databases** (Brown University, CSC2270) – Spring 2011

Mentoring

PhD Students

- Sam Arch (2022 – present)
- William Zhang (2021 – present)
- Wan Shen Lim (2020 – present)
- Matt Butrovich (2018 – 2024)
- Lin Ma (2015 – 2021)
- Prashanth Menon (2015 – 2021)
- Dana Van Aken (2014 – 2021)
- Huanchen Zhang (2014 – 2019)
- Joy Arulraj (2013 – 2018)
- Xiangyao Yu (2013 – 2017)

Post-Docs

- Lin Ma (2021 – 2022)
- Andrew Crotty (2021 – 2022)

Master's Students

- Abi Kim (Carnegie Mellon University, B.S., M.Sc. 2023)
- Mike Xu (Carnegie Mellon University, B.S., M.Sc. 2023)
- Deepayan Patra (Carnegie Mellon University, B.S., M.Sc. 2022)
- Tanuj Nayak (Carnegie Mellon University, B.S., M.Sc. 2021)
- Amadou Ngom (Carnegie Mellon University, B.S., M.Sc. 2020)
- Lily Liu (Carnegie Mellon University, M.Sc. 2019)
- Gustavo Angulo (Carnegie Mellon University, B.S., M.Sc. 2019)

- Tianyu Li (Carnegie Mellon University, B.S., M.Sc. 2019)
- Sivaprasad Sudhir (Carnegie Mellon University, M.Sc. 2018)
- Prashasthi Prabhakar (Carnegie Mellon University, M.Sc. 2017)
- Patrick Huang (Carnegie Mellon University, M.Sc. 2017)
- Matt Perron (Carnegie Mellon University, M.Sc. 2016)
- Thomas Marshall (Carnegie Mellon University, M.Sc. 2015)
- Atreyee Maiti (Carnegie Mellon University, M.Sc. 2014)
- Xin Jia (Brown University, M.Sc. 2013)
- Yang Zou (Brown University, M.Sc. 2012)
- Yang Lu (Brown University, M.Sc. 2012)
- Visawee Angkanawaraphan (Brown University, M.Sc. 2011)
- Zhe Zhang (Brown University, M.Sc. 2010)
- Ning Shi (Brown University, M.Sc. 2010)

Awards & Honors

- VLDB Early Career Award, Summer 2021
- ACM SIGMOD Distinguished Program Committee Member, Summer 2019
- NSF CAREER, Winter 2019
- VLDB Distinguished Associate Editor, Summer 2018
- ACM SIGMOD Best Paper Award, Summer 2018
- ACM SIGMOD Distinguished Program Committee Member, Summer 2018
- CMU Joel & Ruth Spira Teaching Award, Spring 2018
- Sloan Research Fellowship, Winter 2018
- Google Faculty Research Award, Winter 2018
- ACM SIGMOD Distinguished Program Committee Member, Summer 2017
- Facebook Faculty Research Award, Summer 2015
- Google Faculty Research Award, Summer 2015
- ACM SIGMOD Jim Gray Best Dissertation Award, Summer 2014
- Amazon Research Grant, Winter 2012
- ACM SIGMOD Student Travel Grant, Spring 2010
- Class Clown, Mt. Hebron High School, Spring 1999

Service

To the University

- CSD Hiring Committee – 2022
- SCS Undergraduate Curriculum Review Committee – 2020-present
- CSD Hiring Committee – 2018
- CSD MS Admissions Committee – 2017
- CSD PhD Admissions Committee – 2016
- CSD PhD Admissions Committee – 2015
- CSD PhD Admissions Committee – 2014

To the Profession

- Program Committee / Board of Trustees – CIDR (2021-present)
- Editorial Board – Foundations and Trends in Databases (2020-2022)
- Program Committee – CIDR 2020
- Associate Editor – VLDB 2020
- Area Chair – SIGMOD 2020
- Program Committee – CIDR 2019
- Program Committee – SIGMOD 2019
- Program Committee – VLDB 2019
- Associate Editor – VLDB 2018
- Program Committee – SIGMOD 2018

- Co-Chair – FADS@VLDB 2017
- Program Committee / Demo Co-Chair – SIGMOD 2017
- Program Committee – CIDR 2017
- Program Committee Area Chair – ICDE 2016
- Program Committee / PhD Workshop – VLDB 2016
- Co-Chair – IMDM 2016
- Program Committee – SIGMOD 2016
- Steering Committee / Program Committee – HPTS 2015
- Co-Chair – IMDM 2015
- Program Committee / Demo Committee / Dissertation Award Committee – SIGMOD 2015
- Program Committee / Demo Committee – VLDB 2015
- Co-Chair – IMDM 2014
- Program Committee – SIGMOD 2014
- Program Committee – SoCC 2014
- Program Committee – CloudDP 2014
- Program Committee – VLDB 2014
- External Reviewer – LNCS Transactions on Large-Scale Data- and Knowledge-Centered Systems, March 2013
- External Reviewer – Springer Journal of Distributed and Parallel Databases, January 2013
- External Reviewer – IBM Journal of Research and Development, July 2012
- Demo Judge – SIGMOD 2010
- Demo Judge – SIGMOD 2009

Publications

Refereed

- [1] Xinyu Zeng, Ruijun Meng, Andrew Pavlo, Wes McKinney, and Huanchen Zhang. “NULLS!: Revisiting Null Representation in Modern Columnar Formats”. In: *Proceedings of the 20th International Workshop on Data Management on New Hardware*. DaMoN ’24. 2024. DOI: 10.1145/3662010.3663452. URL: <https://db.cs.cmu.edu/papers/2024/zeng-damon24.pdf>.
- [2] Kai Franz, Samuel I Arch, Denis Hirn, Torsten Grust, Todd Mowry, and Andrew Pavlo. “Dear User-Defined Functions, Inlining isn’t working out so great for us. Let’s try batching to make our relationship work. Sincerely, SQL”. In: *CIDR 2024, Conference on Innovative Data Systems Research*. 2024. URL: <https://db.cs.cmu.edu/papers/2024/p13-franz.pdf>.
- [3] Xinyu Zeng, Yulong Hui, Jiahong Shen, Andrew Pavlo, Wes McKinney, and Huanchen Zhang. “An Empirical Evaluation of Columnar Storage Formats”. In: *Proc. VLDB Endow.* 17.2 (2023), pp. 148–161. URL: <https://www.vldb.org/pvldb/vol17/p148-zeng.pdf>.
- [4] Matthew Butrovich, Karthik Ramanathan, John Rollinson, Wan Shen Lim, William Zhang, Justine Sherry, and Andrew Pavlo. “Tigger: A Database Proxy That Bounces With User-Bypass”. In: *Proc. VLDB Endow.* 16.11 (2023), pp. 3335–3348. DOI: 10.14778/3611479.3611530. URL: <https://www.vldb.org/pvldb/vol16/p3335-butrovich.pdf>.
- [5] Wan Shen Lim, Matthew Butrovich, William Zhang, Andrew Crotty, Lin Ma, Peijing Xu, Johannes Gehrke, and Andrew Pavlo. “Database Gyms”. In: *CIDR 2023, Conference on Innovative Data Systems Research*. 2023. URL: <https://db.cs.cmu.edu/papers/2023/p27-lim.pdf>.
- [6] Yu Xia, Xiangyao Yu, Matthew Butrovich, Andrew Pavlo, and Srinivas Devadas. “Litmus: Towards a Practical Database Management System with Verifiable ACID Properties and Transaction Correctness”. In: *Proceedings of the 2022 International Conference on Management of Data*. SIGMOD ’22. 2022, pp. 1478–1492. DOI: 10.1145/3514221.3517851. URL: <https://db.cs.cmu.edu/papers/2022/yu-litmusdb-sigmod2022.pdf>.
- [7] Matthew Butrovich, Wan Shen Lim, Lin Ma, John Rollinson, William Zhang, Yu Xia, and Andrew Pavlo. “Tastes Great! Less Filling! High Performance and Accurate Training Data Collection for Self-Driving Database Management Systems”. In: *Proceedings of the 2022 International Conference on Management of Data*. SIGMOD/PODS ’22. 2022, pp. 617–630. DOI: 10.1145/3514221.3517845.

- [8] Andrew Crotty, Viktor Leis, and Andrew Pavlo. “Are You Sure You Want to Use MMAP in Your Database Management System?” In: *CIDR 2022, Conference on Innovative Data Systems Research*. 2022. URL: <https://db.cs.cmu.edu/papers/2022/cidr2022-p13-crotty.pdf>.
- [9] Dana Van Aken, Dongsheng Yang, Sebastien Brillard, Ari Fiorino, Bohan Zhang, Christian Billian, and Andrew Pavlo. “An Inquiry into Machine Learning-based Automatic Configuration Tuning Services on Real-World Database Management Systems”. In: *Proc. VLDB Endow.* 14.7 (2021), pp. 1241–1253. URL: <https://db.cs.cmu.edu/papers/2021/p1241-aken.pdf>.
- [10] Lin Ma, William Zhang, Jie Jiao, Wuwen Wang, Matthew Butrovich, Wan Shen Lim, Prashanth Menon, and Andrew Pavlo. “MB2: Decomposed Behavior Modeling for Self-Driving Database Management Systems”. In: *Proceedings of the 2021 International Conference on Management of Data*. SIGMOD ’21. 2021, pp. 1248–1261. DOI: 10.1145/3448016.3457276. URL: <https://db.cs.cmu.edu/papers/2021/ma-sigmod2021.pdf>.
- [11] Amadou Ngom, Prashanth Menon, Matthew Butrovich, Lin Ma, Wan Shen Lim, Todd C. Mowry, and Andrew Pavlo. “Filter Representation in Vectorized Query Execution”. In: *Proceedings of the 17th International Workshop on Data Management on New Hardware (DaMoN 2021)*. DAMON’21. 2021, 6:1–6:7. DOI: 10.1145/3465998.3466009. URL: <https://db.cs.cmu.edu/papers/2021/ngom-damon2021.pdf>.
- [12] Xinjing Zhou, Joy Arulraj, Andrew Pavlo, and David Cohen. “Spitfire: A Three-Tier Buffer Manager for Volatile and Non-Volatile Memory”. In: *Proceedings of the 2021 International Conference on Management of Data*. SIGMOD ’21. 2021, pp. 2195–2207. DOI: 10.1145/3448016.3452819. URL: <https://db.cs.cmu.edu/papers/2021/zhou-sigmod2021.pdf>.
- [13] Ling Zhang, Matthew Butrovich, Tianyu Li, Andrew Pavlo, Yash Nannapaneni, John Rollinson, Huanchen Zhang, Ambarish Balakumar, Daniel Biales, Ziqi Dong, Emmanuel J Eppinger, Jordi E Gonzalez, Wan Shen Lim, Jianqiao Liu, Lin Ma, Prashanth Menon, Soumil Mukherjee, Tanuj Nayak, Amadou Ngom, Dong Niu, Deepayan Patra, Poojita Raj, Stephanie Wang, Wuwen Wang, Yao Yu, and William Zhang. “Everything is a Transaction: Unifying Logical Concurrency Control and Physical Data Structure Maintenance in Database Management Systems”. In: *CIDR 2021, Conference on Innovative Data Systems Research*. 2021. URL: https://db.cs.cmu.edu/papers/2021/cidr2021_paper06.pdf.
- [14] Tianyu Li, Matthew Butrovich, Amadou Ngom, Wan Shen Lim, Wes McKinney, and Andrew Pavlo. “Mainlining Databases: Supporting Fast Transactional Workloads on Universal Columnar Data File Formats”. In: *Proc. VLDB Endow.* 14.4 (2020), pp. 534–546. URL: <https://db.cs.cmu.edu/papers/2020/p534-li.pdf>.
- [15] Yu Xia, Xiangyao Yu, Andrew Pavlo, and Srinivas Devadas. “Taurus: Lightweight Parallel Logging for In-Memory Database Management Systems”. In: *Proc. VLDB Endow.* 14.2 (2020), pp. 189–201. URL: <https://db.cs.cmu.edu/papers/2020/p189-xia.pdf>.
- [16] Prashanth Menon, Amadou Ngom, and Andrew Pavlo Lin Ma Todd C. Mowry. “Permutable Compiled Queries: Dynamically Adapting Compiled Queries without Recompiling”. In: *Proc. VLDB Endow.* 14.2 (2020), pp. 101–113. URL: <https://db.cs.cmu.edu/papers/2020/p101-memon.pdf>.
- [17] Huanchen Zhang, Xiaoxuan Liu, David G. Andersen, Michael Kaminsky, Kimberly Keeton, and Andrew Pavlo. “Order-Preserving Key Compression for In-Memory Search Trees”. In: *Proceedings of the 2020 International Conference on Management of Data*. 2020, pp. 1601–1615. URL: <https://db.cs.cmu.edu/papers/2020/zhang-sigmod2020.pdf>.
- [18] Yihan Sun, Guy E. Blelloch, Wan Shen Lim, and Andrew Pavlo. “On Supporting Efficient Snapshot Isolation for Hybrid Workloads with Multi-Versioned Indexes”. In: *Proc. VLDB Endow.* 13 (2 Oct. 2019), pp. 221–225. URL: <https://db.cs.cmu.edu/papers/2019/p211-sun.pdf>.
- [19] Yangjun Sheng, Anthony Tomasic, Tieying Zhang, and Andrew Pavlo. “Scheduling OLTP transactions via learned abort prediction”. In: *Proceedings of the Second International Workshop on Exploiting Artificial Intelligence Techniques for Data Management, aiDM@SIGMOD 2019*. 2019, 1:1–1:8. URL: <https://db.cs.cmu.edu/papers/2019/a1-sheng.pdf>.
- [20] Timo Kersten, Viktor Leis, Alfons Kemper, Thomas Neumann, Andrew Pavlo, and Peter Boncz. “Everything You Always Wanted to Know About Compiled and Vectorized Queries But Were Afraid to Ask”. In: *Proc. VLDB Endow.* 11 (13 Sept. 2018), pp. 2209–2222. URL: <https://db.cs.cmu.edu/papers/2018/p2209-kersten.pdf>.
- [21] Xiangyao Yu, Yu Xia, Andrew Pavlo, Daniel Sanchez, Larry Rudolph, and Srinivas Devadas. “Sundial: Harmonizing Concurrency Control and Caching in a Distributed OLTP Database Management System”. In: *Proc. VLDB Endow.* 11.10 (June 2018), pp. 1289–1302. URL: <https://db.cs.cmu.edu/papers/2018/p1289-yu.pdf>.

- [22] Ziqi Wang, Andrew Pavlo, Hyeontaek Lim, Viktor Leis, Huanchen Zhang, Michael Kaminsky, and David G. Andersen. “Building a Bw-Tree Takes More Than Just Buzz Words”. In: *Proceedings of the 2018 ACM International Conference on Management of Data*. SIGMOD ’18. 2018, pp. 473–488. URL: <https://db.cs.cmu.edu/papers/2018/mod342-wangA.pdf>.
- [23] Lin Ma, Dana Van Aken, Ahmed Hefny, Gustavo Mezerhane, Andrew Pavlo, and Geoffrey J Gordon. “Query-based Workload Forecasting for Self-Driving Database Management Systems”. In: *Proceedings of the 2018 ACM International Conference on Management of Data*. SIGMOD ’18. 2018, pp. 631–645. DOI: 10.1145/3183713.3196908. URL: <https://db.cs.cmu.edu/papers/2018/mod435-maA.pdf>.
- [24] Huanchen Zhang, David G. Andersen, Michael Kaminsky, Andrew Pavlo, Hyeontaek Lim, Viktor Leis, and Kimberly Keeton. “SuRF: Practical Range Query Filtering with Fast Succinct Tries”. In: *Proceedings of the 2018 ACM International Conference on Management of Data*. SIGMOD ’18. 2018, pp. 323–336. URL: <https://db.cs.cmu.edu/papers/2018/mod601-zhangA-hm.pdf>.
- [25] Prashanth Menon, Todd C. Mowry, and Andrew Pavlo. “Relaxed Operator Fusion for In-Memory Databases: Making Compilation, Vectorization, and Prefetching Work Together At Last”. In: *Proc. VLDB Endow.* 11.1 (Sept. 2017), pp. 1–13. URL: <https://db.cs.cmu.edu/papers/2017/p1-menon.pdf>.
- [26] Yingjun Wu, Joy Arulraj, Jiexi Lin, Ran Xian, and Andrew Pavlo. “An Empirical Evaluation of In-Memory Multi-Version Concurrency Control”. In: *Proc. VLDB Endow.* 10.7 (Mar. 2017), pp. 781–792. URL: <https://db.cs.cmu.edu/papers/2017/p781-wu.pdf>.
- [27] Rachael Harding, Dana Van Aken, Andrew Pavlo, and Michael Stonebraker. “An Evaluation of Distributed Concurrency Control”. In: *Proc. VLDB Endow.* 10.5 (Jan. 2017), pp. 553–564. URL: <https://db.cs.cmu.edu/papers/2017/p553-harding.pdf>.
- [28] Dana Van Aken, Andrew Pavlo, Geoffrey J. Gordon, and Bohan Zhang. “Automatic Database Management System Tuning Through Large-scale Machine Learning”. In: *Proceedings of the 2017 ACM International Conference on Management of Data*. SIGMOD ’17. 2017, pp. 1009–1024. DOI: 10.1145/3035918.3064029. URL: <https://db.cs.cmu.edu/papers/2017/p1009-van-aken.pdf>.
- [29] Lianghong Xu, Andrew Pavlo, Sudipta Sengupta, and Gregory R. Ganger. “Online Deduplication for Databases”. In: *Proceedings of the 2017 ACM International Conference on Management of Data*. SIGMOD ’17. 2017, pp. 1355–1368. URL: <https://db.cs.cmu.edu/papers/2017/p1355-xu.pdf>.
- [30] Andrew Pavlo, Gustavo Angulo, Joy Arulraj, Haibin Lin, Jiexi Lin, Lin Ma, Prashanth Menon, Todd Mowry, Matthew Perron, Ian Quah, Siddharth Santurkar, Anthony Tomasic, Skye Toor, Dana Van Aken, Ziqi Wang, Yingjun Wu, Ran Xian, and Tieying Zhang. “Self-Driving Database Management Systems”. In: *CIDR 2017, Conference on Innovative Data Systems Research*. 2017. URL: <https://db.cs.cmu.edu/papers/2017/p42-pavlo-cidr17.pdf>.
- [31] Joy Arulraj, Matthew Perron, and Andrew Pavlo. “Write-Behind Logging”. In: *Proc. VLDB Endow.* 10.4 (Dec. 2016), pp. 337–348. URL: <https://db.cs.cmu.edu/papers/2016/p337-arulraj.pdf>.
- [32] Marco Serafini, Rebecca Taft, Aaron J Elmore, Andrew Pavlo, Ashraf Aboulmaga, and Michael Stonebraker. “Clay: Fine-Grained Adaptive Partitioning for General Database Schemas”. In: *Proc. VLDB Endow.* 10.4 (Dec. 2016), pp. 445–456. URL: <http://hstore.cs.brown.edu/papers/hstore-clay.pdf>.
- [33] Lin Ma, Joy Arulraj, Sam Zhao, Andrew Pavlo, Subramanya R. Dullloor, Michael J. Giardino, Jeff Parkhurst, Jason L. Gardner, Kshitij Doshi, and Stanley Zdonik. “Larger-than-memory Data Management on Modern Storage Hardware for In-memory OLTP Database Systems”. In: *Proceedings of the 12th International Workshop on Data Management on New Hardware*. DaMoN ’16. 2016, 9:1–9:7. URL: <https://db.cs.cmu.edu/papers/2016/madamon2016.pdf>.
- [34] Joy Arulraj, Andrew Pavlo, and Prashanth Menon. “Bridging the Archipelago Between Row-Stores and Column-Stores for Hybrid Workloads”. In: *Proceedings of the 2016 International Conference on Management of Data*. SIGMOD ’16. 2016, pp. 583–598. DOI: 10.1145/2882903.2915231. URL: <https://db.cs.cmu.edu/papers/2016/arulraj-sigmod2016.pdf>.
- [35] Huanchen Zhang, David G. Andersen, Andrew Pavlo, Michael Kaminsky, Lin Ma, and Rui Shen. “Reducing the Storage Overhead of Main-Memory OLTP Databases with Hybrid Indexes”. In: *Proceedings of the 2016 International Conference on Management of Data*. SIGMOD ’16. 2016, pp. 1567–1581. URL: <https://db.cs.cmu.edu/papers/2016/zhang-sigmod2016.pdf>.

- [36] Xiangyao Yu, Andrew Pavlo, Daniel Sanchez, and Srinivas Devadas. “TicToc: Time Traveling Optimistic Concurrency Control”. In: *Proceedings of the 2016 International Conference on Management of Data*. SIGMOD ’16. 2016, pp. 1629–1642. DOI: 10.1145/2882903.2882935. URL: <https://db.cs.cmu.edu/papers/2016/yu-sigmod2016.pdf>.
- [37] John Meehan, Nesime Tatbul, Stan Zdonik, Cansu Aslantas, Ugur Çetintemel, Jiang Du, Samuel Madden, David Maier, Andrew Pavlo, Michael Stonebraker, Kristin Tufte, and Hao Wang. “S-Store: Streaming Meets Transaction Processing”. In: *PVLDB 8.13* (2015), pp. 2134–2145. URL: <https://db.cs.cmu.edu/papers/2015/p2134-meehan.pdf>.
- [38] Lianghong Xu, Andrew Pavlo, Sudipta Sengupta, Jin Li, and Gregory R. Ganger. “Reducing Replication Bandwidth for Distributed Document Databases”. In: *Proceedings of the Sixth ACM Symposium on Cloud Computing*. SoCC ’15. 2015, pp. 222–235. URL: <https://db.cs.cmu.edu/papers/2015/socc15-sdedup.pdf>.
- [39] Aaron J. Elmore, Vaibhav Arora, Rebecca Taft, Andrew Pavlo, Divyakant Agrawal, and Amr El Abbadi. “Squall: Fine-Grained Live Reconfiguration for Partitioned Main Memory Databases”. In: *Proceedings of SIGMOD*. SIGMOD ’15. 2015, pp. 299–313. URL: <https://db.cs.cmu.edu/papers/2015/p299-elmore.pdf>.
- [40] Joy Arulraj, Andrew Pavlo, and Subramanya Dullloor. “Let’s Talk About Storage & Recovery Methods for Non-Volatile Memory Database Systems”. In: *Proceedings of the 2015 International Conference on Management of Data*. SIGMOD ’15. 2015, pp. 707–722. URL: <https://db.cs.cmu.edu/papers/2015/p707-arulraj.pdf>.
- [41] Xiangyao Yu, George Bezerra, Andrew Pavlo, Srinivas Devadas, and Michael Stonebraker. “Staring into the Abyss: An Evaluation of Concurrency Control with One Thousand Cores”. In: *PVLDB 8.3* (2014), pp. 209–220. URL: <https://www.vldb.org/pvldb/vol8/p209-yu.pdf>.
- [42] Rebecca Taft, Essam Mansour, Marco Serafini, Jennie Duggan, Aaron J. Elmore, Ashraf Aboulnaga, Andrew Pavlo, and Michael Stonebraker. “E-Store: Fine-Grained Elastic Partitioning for Distributed Transaction Processing”. In: *PVLDB 8.3* (2014), pp. 245–256. URL: <https://hstore.cs.brown.edu/papers/hstore-elastic.pdf>.
- [43] Justin DeBrabant, Joy Arulraj, Andrew Pavlo, Michael Stonebraker, Stanley B. Zdonik, and Subramanya Dullloor. “A Prolegomenon on OLTP Database Systems for Non-Volatile Memory”. In: *ADMS @ VLDB*. 2014, pp. 57–63. URL: <http://hstore.cs.brown.edu/papers/hstore-nvm.pdf>.
- [44] Djellel Eddine Difallah, Andrew Pavlo, Carlo Curino, and Philippe Cudré-Mauroux. “OLTP-Bench: An Extensible Testbed for Benchmarking Relational Databases”. In: *PVLDB 7.4* (2013), pp. 277–288. URL: <https://www.cs.cmu.edu/~pavlo/static/papers/oltpbench-vldb.pdf>.
- [45] Justin DeBrabant, Andrew Pavlo, Stephen Tu, Michael Stonebraker, and Stanley B. Zdonik. “Anti-Caching: A New Approach to Database Management System Architecture”. In: *PVLDB 6.14* (2013), pp. 1942–1953. URL: <http://hstore.cs.brown.edu/papers/hstore-anticaching.pdf>.
- [46] Carlo A. Curino, Djellel E. Difallah, Andrew Pavlo, and Philippe Cudre-Mauroux. “Benchmarking OLTP/Web Databases in the Cloud: The OLTP-bench Framework”. In: *CloudDB*. 2012, pp. 17–20. URL: <http://www.cs.cmu.edu/~pavlo/static/papers/oltpbench.pdf>.
- [47] Andrew Pavlo, Carlo Curino, and Stanley Zdonik. “Skew-Aware Automatic Database Partitioning in Shared-Nothing, Parallel OLTP Systems”. In: *SIGMOD*. 2012, pp. 61–72. URL: <https://hstore.cs.brown.edu/papers/hstore-partitioning.pdf>.
- [48] Andrew Pavlo, Evan P.C. Jones, and Stan Zdonik. “On Predictive Modeling for Optimizing Transaction Execution in Parallel OLTP Systems”. In: *Proc. VLDB Endow.* 5 (2 Oct. 2011), pp. 85–96. URL: <https://hstore.cs.brown.edu/papers/hstore-markov.pdf>.
- [49] Andrew Pavlo, Erik Paulson, Alexander Rasin, Daniel J. Abadi, David J. DeWitt, Samuel Madden, and Michael Stonebraker. “A Comparison of Approaches to Large-Scale Data Analysis”. In: *Proceedings of SIGMOD*. 2009, pp. 165–178. URL: <https://www.cs.cmu.edu/~pavlo/papers/benchmarks-sigmod09.pdf>.
- [50] Ewa Deelman, Miron Livny, Gaurang Mehta, Andy Pavlo, Gurmeet Singh, Mei-Hui Su, Karan Vahi, and R. Kent Wenger. “Pegasus and DAGMan From Concept to Execution: Mapping Scientific Workflows onto Today’s Cyberinfrastructure”. In: *HPC and Grids in Action*. Vol. 16. Advances in Parallel Computing. 2008, pp. 56–74. URL: https://www.cs.cmu.edu/~pavlo/static/papers/cetraro_final.pdf.
- [51] Andrew Pavlo, Peter Couvares, Rebekah Gietzel, Anatoly Karp, Ian D. Alderman, Miron Livny, and Charles Bacon. “The NMI Build & Test Laboratory: Continuous Integration Framework for Distributed Computing Software”. In: *LISA*. 2006, pp. 263–273. URL: <https://www.cs.cmu.edu/~pavlo/static/papers/nmi-lisa2006.pdf>.

Books

- [1] Joy Arulraj and Andrew Pavlo. *Non-Volatile Memory Database Management Systems*. Synthesis Lectures on Data Management. Morgan & Claypool Publishers, 2019. doi: 10.2200/S00891ED1V01Y201812DTM055.

Book Chapters

- [1] Andy Pavlo. “Making Databases Work”. In: ed. by Michael L. Brodie. Association for Computing Machinery and Morgan & Claypool, 2019. Chap. In-memory, Horizontal, and Transactional: The H-store OLTP DBMS Project, pp. 245–251.
- [2] Franz Faerber, Alfons Kemper, Per-Åke Larson, Justin J. Levandoski, Thomas Neumann, and Andrew Pavlo. “Main Memory Database Systems”. In: *Foundations and Trends in Databases* 8.1-2 (2017), pp. 1–130.

Demonstrations / Tutorials

- [1] Bohan Zhang, Dana Van Aken, Justin Wang, Tao Dai, Shuli Jiang, Jacky Lao, Siyuan Sheng, Andrew Pavlo, and Geoffrey J. Gordon. “A Demonstration of the OtterTune Automatic Database Management System Tuning Service”. In: *PVLDB* 11.12 (2018), pp. 1910–1913.
- [2] Joy Arulraj and Andrew Pavlo. “How to Build a Non-Volatile Memory Database Management System”. In: *Proceedings of the 2017 ACM International Conference on Management of Data*. SIGMOD ’17. 2017, pp. 1753–1758.
- [3] Dana Van Aken, Djelle Eddine Difallah, Andrew Pavlo, Carlo Curino, and Philippe Cudre-Mauroux. “BenchPress: Dynamic Workload Control in the OLTP-Bench Testbed”. In: *Proceedings of SIGMOD*. 2015, pp. 1069–1073.
- [4] Ugur Çetintemel, Jiang Du, Tim Kraska, Samuel Madden, David Maier, John Meehan, Andrew Pavlo, Michael Stonebraker, Erik Sutherland, Nesime Tatbul, Kristin Tufte, Hao Wang, and Stanley B. Zdonik. “S-Store: A Streaming NewSQL System for Big Velocity Applications”. In: *PVLDB* 7.13 (2014), pp. 1633–1636.
- [5] Robert Kallman, Hideaki Kimura, Jonathan Natkins, Andrew Pavlo, Alexander Rasin, Stanley Zdonik, Evan P. C. Jones, Samuel Madden, Michael Stonebraker, Yang Zhang, John Hugg, and Daniel J. Abadi. “H-Store: A High-Performance, Distributed Main Memory Transaction Processing System”. In: *Proc. VLDB Endow.* 1.2 (2008), pp. 1496–1499.
- [6] Christopher Homan, Andrew Pavlo, and Jonathan Schull. “Smoother Transitions Between Breadth-First-Spanning-Tree-Based Drawings”. In: *14th International Symposium on Graph Drawing*. Vol. 4372. Lecture Notes in Computer Science. 2006, pp. 442–445.

Miscellaneous Writings

- [1] Huanchen Zhang, Hyeontaek Lim, Viktor Leis, David G. Andersen, Michael Kaminsky, Kimberly Keeton, and Andrew Pavlo. “Succinct Range Filters”. In: *Commun. ACM* 64.4 (Mar. 2021), pp. 166–173. URL: <https://doi.org/10.1145/3450262>.
- [2] Andrew Pavlo, Matthew Butrovich, Lin Ma, Prashanth Menon, Wan Shen Lim, Dana Van Aken, and William Zhang. “Make Your Database System Dream of Electric Sheep: Towards Self-Driving Operation”. In: *Proc. VLDB Endow.* 14.12 (2021), pp. 3211–3221. URL: <http://www.vldb.org/pvldb/vol14/p3211-pavlo.pdf>.
- [3] Huanchen Zhang, Hyeontaek Lim, Viktor Leis, David G. Andersen, Michael Kaminsky, Kimberly Keeton, and Andrew Pavlo. “Succinct Range Filters”. In: *ACM Trans. Database Syst.* 45.2 (2020), 5:1–5:31.
- [4] Andrew Pavlo, Matthew Butrovich, Ananya Joshi, Lin Ma, Prashanth Menon, Dana Van Aken, Lisa Lee, and Ruslan Salakhutdinov. “External vs. Internal: An Essay on Machine Learning Agents for Autonomous Database Management Systems”. In: *IEEE Data Engineering Bulletin* (June 2019), pp. 32–46.
- [5] Huanchen Zhang, Hyeontaek Lim, Viktor Leis, David G. Andersen, Michael Kaminsky, Kimberly Keeton, and Andrew Pavlo. “Succinct Range Filters”. In: *SIGMOD Record* 48.1 (Mar. 2019), pp. 78–85.
- [6] Dana Van Aken, Andrew Pavlo, and Geoffrey J. Gordon. *Amazon AWS AI Blog – Tuning Your DBMS Automatically with Machine Learning*. <https://aws.amazon.com/blogs/ai/tuning-your-dbms-automatically-with-machine-learning/>. June 2017.

- [7] Andrew Pavlo. “What Are We Doing With Our Lives?: Nobody Cares About Our Concurrency Control Research”. In: *Proceedings of the 2017 ACM International Conference on Management of Data*. SIGMOD '17. 2017, p. 3.
- [8] Peter Bailis, Camille Fournier, Joy Arulraj, and Andrew Pavlo. “Research for Practice: Distributed Consensus and Implications of NVM on Database Management Systems”. In: *Commun. ACM* 59.11 (Oct. 2016), pp. 52–55. URL: <https://db.cs.cmu.edu/papers/2016/cacm-nvm2016.pdf>.
- [9] Andrew Pavlo and Matthew Aslett. “What’s Really New with NewSQL?” In: *SIGMOD Rec.* 45.2 (June 2016), pp. 45–55. URL: <https://db.cs.cmu.edu/papers/2016/pavlo-newsq-sigmodrec2016.pdf>.
- [10] Andrew Pavlo. “Emerging Hardware Trends in Large-Scale Transaction Processing”. In: *IEEE Internet Computing* 19.3 (May 2015), pp. 68–71. URL: https://www.cs.cmu.edu/~pavlo/static/papers/IC_20150501_May_2015.pdf.
- [11] Michael Stonebraker, Andrew Pavlo, Rebecca Taft, and Michael L. Brodie. “Enterprise Database Applications and the Cloud: A Difficult Road Ahead”. In: *2014 IEEE International Conference on Cloud Engineering*. 2014, pp. 1–6.
- [12] Andrew Pavlo. “Not Your Traditional Data Management - HPTS Conference Report”. In: *USENIX ;login:* (Feb. 2012), pp. 76–78. URL: <https://www.usenix.org/publications/login/february-2012/14th-international-workshop-high-performance%20-transaction-systems>.
- [13] Xin Jia, Andrew Pavlo, and Stanley B. Zdonik. “Tastes Great, Less Filling: Low-Impact OLAP MapReduce Queries on High-Performance OLTP Systems”. In: *TinyToCS* 1 (2012). URL: <http://tinytocs.org/vol1/papers/tinytocs-v1-jia.pdf>.
- [14] Andrew Pavlo and Ning Shi. “Graffiti Networks: A Subversive, Internet-Scale File Sharing Model”. In: *CoRR* abs/1101.0350 (2011). URL: <http://arxiv.org/abs/1101.0350>.
- [15] Michael Stonebraker, Daniel Abadi, David J. DeWitt, Sam Madden, Erik Paulson, Andrew Pavlo, and Alexander Rasin. “MapReduce and Parallel DBMSs: Friends or Foes?” In: *Communications of the ACM* 53.1 (2010), pp. 64–71.
- [16] Andrew Pavlo, Christopher Homan, and Jonathan Schull. “A parent-centered radial layout algorithm for interactive graph visualization and animation”. In: *CoRR* abs/cs/0606007 (2006). URL: <http://arxiv.org/abs/cs/0606007>.

Synergistic Activities

Open-Source Software:

- BenchBase Database Benchmark Suite (<https://github.com/oltpbenchmark>)
- NoisePage Database Management System (<https://noisepage.com>)
- Peloton Database Management System (<https://pelotondb.io>)
- OtterTune (<https://ottertune.cs.cmu.edu>)
- Database of Databases (<https://dbdb.io>)
- Carnegie Mellon Database Application Catalog (<http://cmdbac.cs.cmu.edu>)
- H-Store Database Management System (<https://hstore.cs.brown.edu>)

Database Seminar Series:

- Fall 2023: ML \rightleftharpoons DB Technical Talks (<https://db.cs.cmu.edu/seminar2023>)
- Fall 2022: ¡Databases! - A Database Seminar Series (<https://db.cs.cmu.edu/seminar2022>)
- Spring 2022: Vaccination Database Tech Talks – Booster (<https://db.cs.cmu.edu/seminar2022-booster>)
- Fall 2021: Vaccination Database Tech Talks – Second Dose (<https://db.cs.cmu.edu/seminar2021-dose2>)
- Spring 2021: Vaccination Database Tech Talks – First Dose (<https://db.cs.cmu.edu/seminar2021>)
- Fall 2020: Quarantine Database Tech Talks (<https://db.cs.cmu.edu/seminar2020>)
- Fall 2018: Hardware-Accelerated Database Lectures (<https://db.cs.cmu.edu/seminar2018>)
- Fall 2017: Time Series Database Lectures (<https://db.cs.cmu.edu/seminar2017>)
- Fall 2015: Databaseology Lectures (<https://db.cs.cmu.edu/seminar2015>)
- Fall 2014: Seven Databases in Seven Weeks (<https://db.cs.cmu.edu/seminar2014>)

Internship Programs:

- Summer 2021 (<https://db.cs.cmu.edu/internship2021>)
- Summer 2019 (<https://db.cs.cmu.edu/internship2019>)
- Summer 2018 (<https://db.cs.cmu.edu/internship2018>)
- Summer 2017 (<https://db.cs.cmu.edu/internship2017>)

- Summer 2017 Pennsylvania Governor’s School for the Sciences (Pre-College)

Talks

- “What Goes Around Comes Around... And Around...,” *PGConfNYC 2023 Keynote*, October 2023.
- “Why Machine Learning for Automatically Optimizing Databases Doesn’t Work,” *J on the Beach 2023 Keynote*, May 2023.
- “The Building Blocks for Self-Driving PostgreSQL,” *Citus Con 2022 Keynote*.
- “Do We Still Need People to Write Database Systems?,” *OSACON 2021, Dremio Subsurface 2022*.
- “The OtterTune Database Configuration Auto-Tuning Service,” *University of Iowa / University of Wisconsin / National Database Conference of China, 2020*.
- “Self-Driving Databases & My Pregnant Wife: The Hard Parts,” *Speaking Tour 2019*, <https://db.cs.cmu.edu/tour2019>.
- “Make Your Database Dream of Electric Sheep: Designing for Autonomous Operation,” *Amazon AWS / Percona Live / Microsoft Research Faculty Summit*, March 2018.
- “What Non-Volatile Memory Means for the Future of Database Systems,” *Two Sigma / MongoDB*, August 2017.
- “What Are We Doing With Our Lives?: Nobody Cares About Our Concurrency Control Research,” *SIGMOD Keynote*, May 2017.
- “How to Build a Non-Volatile Memory Database Management System,” *SIGMOD Tutorial*, May 2017.
- “Self-Driving Database Management System,” *CIDR / Google / DataBricks*, January 2017.
- “Peloton: The Self-Driving Database Management System,” *DataEngConf (NYC)*, November 2016.
- “What Non-Volatile Memory Means for the Future of Database Systems,” *In-Memory Computing Summit*, May 2016.
- “Andy Pavlo’s Mostly Professional and Collected Presentation about Getting Down & Dirty in OLTP Databases with Intel’s NVM SDV,” *ISTC for Big Data Retreat*, September 2015.
- “Reducing Replication Bandwidth for Distributed Document Databases,” *MongoDB*, July 2015.
- “Moving the Abyss: Database Management on Future 1000-core Processors,” *NSF Workshop on Exploiting Parallelism and Scalability*, June 2015.
- “I Don’t Want to be the Mitt Romney of Databases,” *University of Michigan / Two Sigma / IBM Research / University of Waterloo*, April–November 2015.
- “Database Systems Do Not Scale to 1000 CPU Cores and Other Tales of the Macabre,” *CMU Qatar / Qatar Computing Research Institute / New England Database Summit*, December 2014.
- “Let’s Talk About Storage & Recovery Methods for Non-Volatile Memory OLTP Database Systems,” *ISTC for Big Data Retreat*, August 2014.
- “H-Store Retrospective,” *SIGMOD Jim Gray Doctoral Dissertation Award*, June 2014.
- “Cache Rules Everything Around Me,” *Univ. of California, Santa Cruz / Univ. of California, Berkeley / Stanford University / University of Pittsburgh*, January 2014
- “OLTP on the NVM SDV: YMMV,” *ISTC for Big Data Retreat*, December 2013.
- “What’s Really New With NewSQL,” *Northwest Database Society / Google Kirkland*, October 2013.
- “Random Prognostications: An Evening with Andy Pavlo,” *HPTS 2013*, September 2013.
- “On Predictive Modeling for Distributed Databases,” *VLDB*, August 2012.
- “Automatic Database Partitioning in Parallel OLTP Systems,” *SIGMOD*, May 2012.
- “Making Fast Databases Faster,” *Columbia University / Yale University*, April 2012.
- “Methods for the Optimization of Parallel OLTP Systems,” *VoltDB, Inc.*, April 2012.

- “Making Fast Databases Faster,” *New England Database Summit 2012*, February 2012.
- “Magical Parallel OLTP Databases,” *Massachusetts Institute of Technology*, November 2011.
- “Life After the Stonebraker Stores,” *HPTS*, October 2011.
- “MapReduce and Parallel DBMSs: Together At Last,” *Brown IPP Symposium on Cloud Computing*, May 2010.
- “MapReduce and Parallel DBMSs: Together At Last,” *New England Database Summit 2010*, January 2010.
- “H-Store: A Specialized Architecture for High-throughput OLTP Applications,” *International Workshop on High Performance Transaction Systems*, October 2009.
- “Graffiti Networks: A Subversive, Internet-Scale File Sharing Model,” *DC401 - Rhode Island Defcon Group*, October 2009.
- “MapReduce and Parallel DBMSs: A Comparison of Approaches to Large-Scale Data Analysis,” *University of Maryland College Park*, September 2009.
- “Challenges in Dynamic Deployment of Condor Across Distributed Environments,” *University of Wisconsin–Madison*, May 2007.
- “NMI Build & Test Laboratory: Continuous Integration Framework for Distributed Computing Software,” *USENIX LISA*, December 2006.

Panels

- “Magical Machine Learning for Database Tuning,” (Moderator) *VLDB*, August 2021.
Video: <https://youtu.be/Z-R8q-LakW8>
- “Junior Faculty Panel,” *CMU/MIT Rising Stars Workshop*, November 2016.
- “Computing Futures,” *CMU CSD50 Celebration*, October 2015.
- “Can Big Data Platforms EVER Deliver Real-time Analytics?,” *BIRTE @ VLDB*, August 2015.
- “Mobility and Big Data: Adding an M to the Vs, or not?,” *IEEE International Conference on Mobile Data Management*, June 2015.